Why a second axis is probably not a good solution

Often we want to compare two different values over time to each other. Since line charts are generally a good way to do this, we may start with such a chart. Sometimes however the numbers are of such a different scale that the whole chart does not make sense, as one of the lines is hardly distinguishable from the x-axis.

A quick solution is to add a second y-axis to the chart. This way we have a nice looking chart, with two lines somewhere in the middle of the picture. And it looks as if we can now easily compare the two data series. But in fact we probably made it even **more difficult** to do so.

An example would be the number of people who visit an online shop, and the number of orders from that shop. The number of visits on a day is typically much larger than the number of orders, and therefore a simple line chart with both data series will show a line for the orders that is very close to the x-axis.

While the solution in this case may be rather obvious, it is still a good example, so let's look at some example data to illustrate the problem.

	Jan	Feb	Mar	Apr	Mai	Jun
# Visits	1000	1500	1600	2000	1900	2100
# Orders	90	100	140	150	170	160

Next we choose four more or less random ways of how to plot this data with two different y-axes. The red line always shows the number of visits, and the blue line always shows the number of orders.
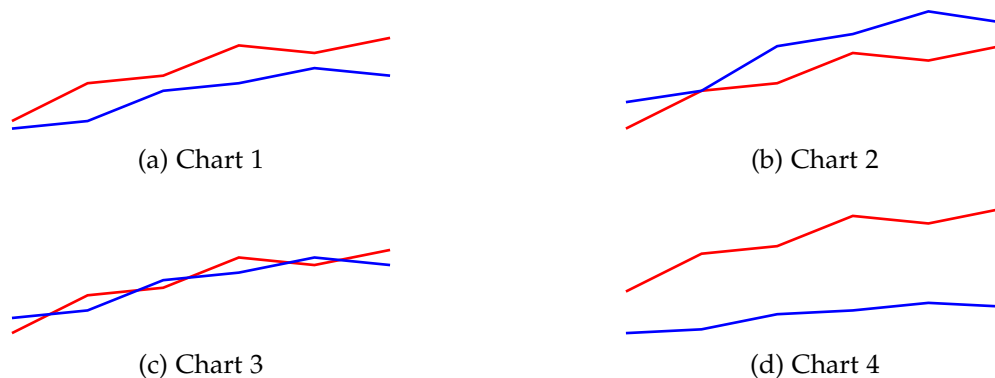

(a) Chart 1

(b) Chart 2

(c) Chart 3

(d) Chart 4

Figure 1: Different ways to plot a line chart with different y-axes.

Notice how it is always the same data! Only the scale of the y-axis for the blue line varies.

- Chart 1 looks as if clicks are "above" orders, but at the end they drift apart.
- Chart 2 looks as if clicks are "below" orders, but at the end they drift closer.
- Chart 3 looks as if the numbers are almost "the same".
- Chart 4 looks as if the number of orders is mostly flat, while visits are increasing.

Now these plots are constructed on purpose to show how different the charts can look (and they are also deliberately missing to display the axes). But there is no way to say one of these four charts (or any other for that matter) is the "correct" one. And all of them will give us a different impression of the data!

The reason why this is a good example, is because drawing a chart like this is actually not what most people do. At least not when they compare visits with orders. What they do instead is to calculate the ratio of them, conveniently called the *conversion rate*.

	Jan	Feb	Mar	Apr	Mai	Jun
Conversion rate	9.00	6.67	8.75	7.50	8.94	7.61

With that ratio, we have one number that clearly tells us how visits and orders compare to each other over time. Plotting the conversion rate gives us a pretty good picture of the situation: the conversion rate is oscillating between a similar range from month to month.
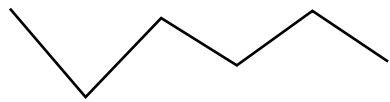
Figure 2: Conversion rate

Not that this solves all our problems. The scale of our y-axis can still make a difference in how we perceive the values. Plotting the same data for the conversion rate with differently scaled y-axes next to each other demonstrates this quite well.
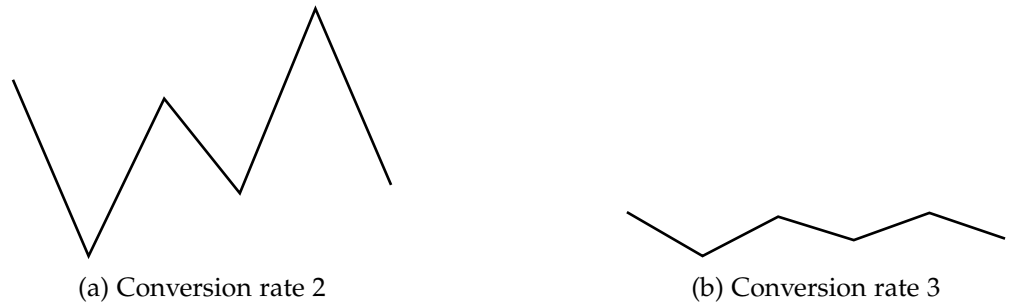
(a) Conversion rate 2   (b) Conversion rate 3

Figure 3: Conversion rates with different scaled y-axes

From the left plot it looks as if there are serious differences from month to month, while the right plot looks rather smooth with only minor variations. Yet all plots on this page show quite clearly that there is a certain zigzag pattern from month to month, which was not really obvious from all (or even any) of the plots before.

Again we can not really say that one plot is "correct" and the other is wrong. But by giving up on a second y-axis, we at least avoided introducing the new problem of how to scale two axis relative to each other. Because finding a reasonable scale for one axis is difficult enough.

Now the solution of using the conversion rate in this example is quite ubiquitous. Nevertheless it demonstrates pretty much exactly the same problem that we nearly always face when two values can not be directly compared. And what the different plots from figure 1 hopefully show, is that just adding a second axis is generally not a very good solution.

A much better solution is to find the "conversion rate" of the problem, which may be the ratio of the numbers to compare, the normalized values of both numbers, or some other number.